Mathematical Biosciences 279 (2016) 83–89

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Identification and automatic segmentation of multiphasic cell growth using a linear hybrid model

András Hartmann[a,*], Ana Rute Neves[b], João M. Lemos[c], Susana Vinga[a]

[a] IDMEC, Instituto Superior Técnico, Universidade de Lisboa - Av. Rovisco Pais, Lisbon 1049-001, Portugal
[b] Discovery, Chr. Hansen A/S - Bøge Alle 10–12, Hørsholm 2970, Denmark
[c] INESC-ID, Instituto Superior Técnico, Universidade de Lisboa - R Alves Redol 9, Lisbon 1000-029, Portugal

ARTICLE INFO

Article history:
Received 16 October 2015
Revised 25 June 2016
Accepted 30 June 2016
Available online 15 July 2016

Keywords:
Hybrid systems
Bacterial growth model
Diauxic growth
Multiphasic growth
System identification

ABSTRACT

This article considers a new mathematical model for the description of multiphasic cell growth. A linear hybrid model is proposed and it is shown that the two-parameter logistic model with switching parameters can be represented by a Switched affine AutoRegressive model with eXogenous inputs (SARX). The growth phases are modeled as continuous processes, while the switches between the phases are considered to be discrete events triggering a change in growth parameters. This framework provides an easily interpretable model, because the intrinsic behavior is the same along all the phases but with a different parameterization. Another advantage of the hybrid model is that it offers a simpler alternative to recent more complex nonlinear models. The growth phases and parameters from datasets of different microorganisms exhibiting multiphasic growth behavior such as *Lactococcus lactis, Streptococcus pneumoniae*, and *Saccharomyces cerevisiae*, were inferred. The segments and parameters obtained from the growth data are close to the ones determined by the experts. The fact that the model could explain the data from three different microorganisms and experiments demonstrates the strength of this modeling approach for multiphasic growth, and presumably other processes consisting of multiple phases.

1. Introduction

Many microorganisms exhibit multiphasic growth behavior. A special case of multiphasic growth caused by substrate preference is diauxic growth, which was first studied in the 40's by Monod [14,15]. This phenomenon arises when an organism is growing on a medium consisting of two (or more) different types of carbon and energy sources. Among others, *Streptococcus pneumoniae* shows diauxic growth on mixed medium. The organism first consumes the substrate that supports the fastest growth (preferred substrate) followed by consumption of the remaining secondary carbon source(s). Theoretically, the process includes two types of phases: exponential growth and diauxic lag. The bacteria process the preferred substrate in an initial exponential growth phase. Then a diauxic lag is followed, when the bacteria do not grow significantly but synthesize enzymes in order to be able to process the next medium in another exponential growth phase. The cycle repeats until no further carbon sources are available.

Yeast (*Saccharomyces cerevisiae*) shows another type of multiphasic growth, where the diauxic behavior is caused by ethanol produced by the fungi itself. The preferred source of carbon and energy of the organism is glucose, but while metabolizing glucose, the cells release ethanol in the medium. When the glucose source becomes limiting, the cells exhibit a so called *diauxic shift* by switching their metabolism to aerobic utilization of ethanol. Compared to glycolysis, the ethanol phase is characterized by decreased growth rate [10].

Inhibitions or metabolic burdens can also cause multiphasic growth. In the *Lactococcus lactis* example considered [16], the biphasic growth behavior can be derived from a metabolic burden due to overexpression of proteins or a metabolic imbalance caused by accumulation of a toxic intermediate.

In order to infer the properties of the process, such as maximal growth, the time-series has to be segmented and a model has to be fitted to each segment. In many cases, segmentation and fitting is still done by hand (usually fitting a linear model to the logarithm of selected points) and checked visually by the experts. Despite tools for automated fitting of biological growth curves like BGFit [22] are readily available online for fitting bacterial growth

* Corresponding author.
E-mail addresses: andras.hartmann@tecnico.ulisboa.pt (A. Hartmann), dkruns@chr-hansen.com (A.R. Neves), jlml@inesc-id.pt (J.M. Lemos), susanavinga@tecnico.ulisboa.pt (S. Vinga).

http://dx.doi.org/10.1016/j.mbs.2016.06.013
0025-5564/

data, they do not consider multiple phases. In this paper a novel approach is considered for multiphasic cell growth segmentation and modeling using hybrid linear systems.

A number of different Ordinary Differential Equation(ODE) models were proposed to describe the growth of biomass on different species starting with the famous hyperbolic Michelis–Menten like equations of [15] to the more sophisticated recent developments. A simple idea to address multiphasic growth is to introduce a lag representing the time-shift between different logistic models [2]. Some approaches include additional state variables incorporating further factors in the model, like enzyme levels or substrate concentrations [5,6,8,20], however this also causes an increased number of parameters, where usually a subset of the parameters is not directly estimated but taken from the literature [5,6].

Hybrid models consist of both continuous and a discrete states (parameters) [13]. The advantage of using hybrid systems is that the dynamics typically can be modeled with continuous state evolution and the transitions between submodels are represented by the changes of discrete states. From strictly computational perspective, it is plausible to interpret the multiphasic growth phenomenon as a hybrid system, because the same model may describe all the phases of the process, while the parameters vary between the phases.

Here a switched hybrid linear model is proposed that is able to explain multiphasic growth data. The model is continuous, and the switches between the phases are modeled with discrete events. Compared to previous models, the advantages of the hybrid model is its linearity, simple and straightforward interpretability, the reduced number of parameters, yet without loss of descriptive power.

2. Methods

In this section, first the proposed model for multiphasic growth is described, then the identification algorithm is reviewed, finally the biological datasets are introduced. A sample MATLAB implementation of the methodology and the proposed model is freely available is freely available under the terms of GNU Public License (GPLv3) from the authors webpage: http://andrashartmann.info/.

The goal of modeling is to accurately describe a given process. Hereafter, a first-order Switched affine AutoRegressive model with eXogenous inputs (SARX) model [17] is proposed for multiphasic growth. SARX models are switched extensions to affine linear autoregressive models, defined as the concatenation of several submodels. Each submodel corresponds to an AutoRegressive model with eXogenous inputs (ARX) of fixed dimension. Consider the following system in input-output form

$$y(t) = \theta_{\eta(t)}^{\mathrm{T}} \begin{bmatrix} \varphi(t) \\ 1 \end{bmatrix} + \varepsilon(t) \tag{1}$$

$$\varphi(t) = [y(t-1) \ldots y(t-n_a)\, u(t) \ldots u(t-n_b)]^{\mathrm{T}}, \tag{2}$$

where the input $u(t)$ is observed and so is the output $y(t)$ which is corrupted by an additional noise term, $\varepsilon(t)$. When fitting the model, $y(t)$ corresponds to the actual observations. The regression vector $\varphi(t)$ of dimension $n = n_a + n_b$ consists of past measurements and inputs. The notation $\Phi(t) = \left[\begin{smallmatrix} \varphi(t) \\ 1 \end{smallmatrix}\right]$ is introduced for the extended regression vector. The time-dependent parameter vector $\theta(t) = \theta_{\eta(t)}^{\mathrm{T}} \in \mathbb{R}^n$ at each time-instance belongs to a set $\Theta = \{\theta_1, \ldots, \theta_K\}$ of cardinality K, representing the submodel set. The discrete finite range function, $\eta(t) : \mathbb{R} \to \{1, \ldots, K\}$ indicates which submodel generates the output at time-instance t, and is referred as switching sequence or discrete state.

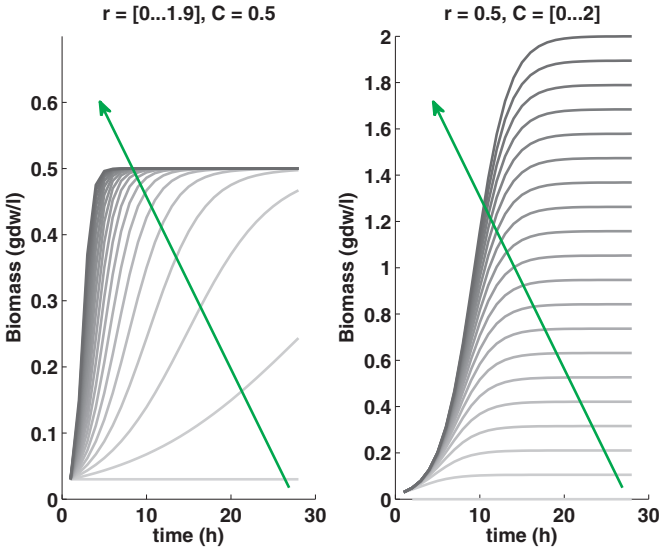

Fig. 1. Simulation example with different parameters. The arrows point towards growing parameters. Also the darker lines represent larger parameter values.

2.1. Logistic growth

As shown in [12], the discrete Verhulst logistic model for monophasic growth [24] has a linear representation, therefore a switching extension of the model can be corresponded to a SARX model. Hereafter, a first-order SARX model is proposed for multiphasic growth by showing that the continuous logistic model also can be represented by an affine AutoRegressive model with eXogenous inputs (ARX). Consequently, when considering switching parameters, the model can be represented by SARX. It is more convenient to work with the continuous model, because no discretization is needed.

The starting point of the model is the continuous two-parameter logistic model.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{C}\right), \tag{3}$$

where the state variable x represents the biomass, the parameter r refers to the maximum growth rate, and C is the carrying capacity (level of saturation). To identify logistic models of longer lag time, some authors [18,24] suggest to fit the logistic model log-measurements $\log(x)$. Here, no log-transformation was applied because no long lag times were expected in the beginning of growth phases. Instead, lag times are considered to be separate phases. The differential equation in the form shown in Eq. (3) is nonlinear. However, since the biomass is strictly positive ($x > 0$), both sides of Eq. (3) can be divided by $x(t)$. Applying the chain rule for the derivation, this also corresponds to the logarithmic derivative

$$\frac{\dot{x}(t)}{x(t)} = r\left(1 - \frac{x(t)}{C}\right) = \frac{\mathrm{d}\log x}{\mathrm{d}t}, \tag{4}$$

with the substitutions of variables

$$y(t) = \frac{\dot{x}(t)}{x(t)}; \quad \theta = \begin{bmatrix} -\frac{r}{C} \\ r \end{bmatrix}, \tag{5}$$

the model may be represented by the following linear autoregressive form

$$y(t) = \theta^{\mathrm{T}} \begin{bmatrix} x(t) \\ 1 \end{bmatrix}. \tag{6}$$

The flexibility of the model is shown in Fig. 1, by simulating with different parameters r and C.

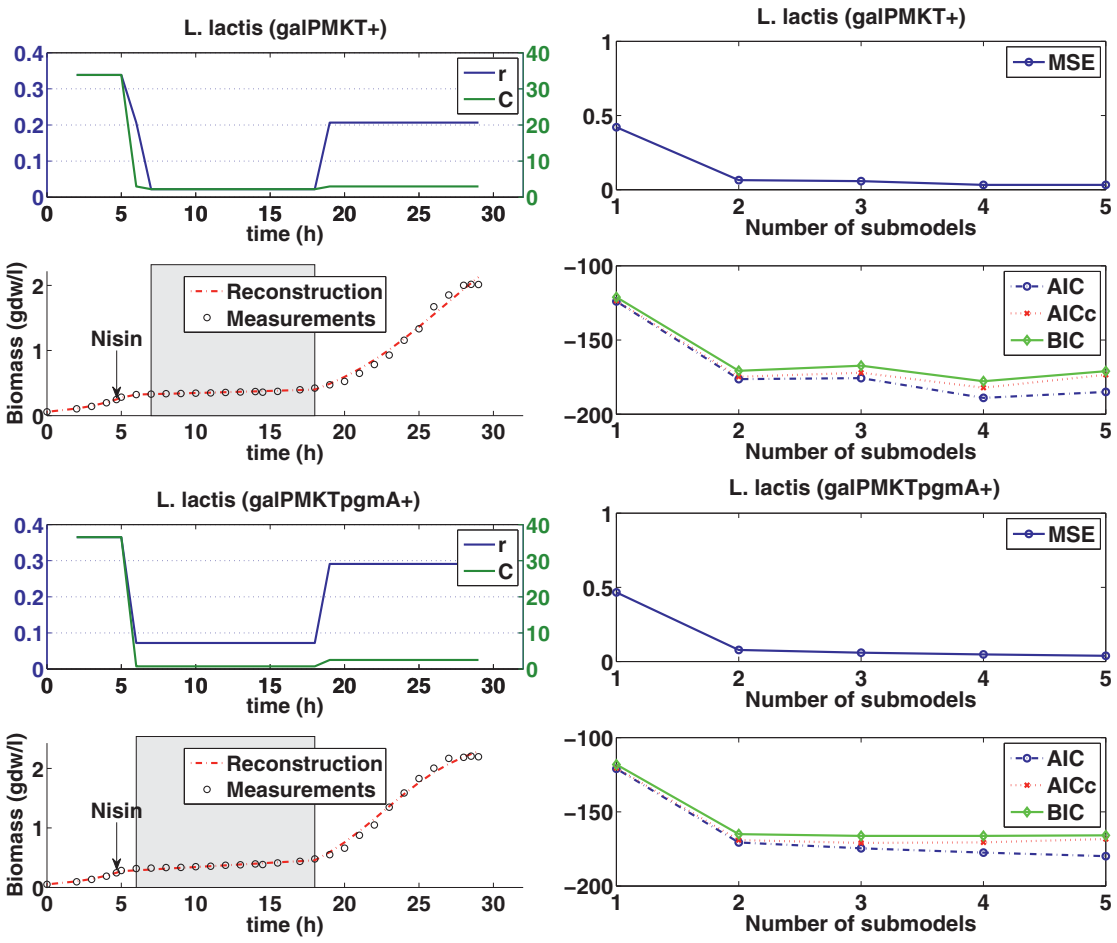

Fig. 2. *Lactococcus lactis* growth on galactose. Expression of galPMKT under the nisin promoter was induced by addition of nisin (1 μg/l) at OD600 of 0.25.

2.2. Model identification

The parameter identification problem for SARX models can be stated as

Problem 1. Given a sequence of input-output pairs $\{u(t), y(t)\}_{t=1}^{T}$, the model orders n_a, n_b and the cardinality of the submodel set K, estimate the parameter vectors $\theta_k;\ k = 1, \ldots, K$ and the switching sequence $\eta(t);\ t = 1, \ldots, T$.

Recent developments on identifying hybrid switched systems provide convenient framework for system identification [11]. The problem of finding the globally optimal parameters of SARX models may be formulated as a mixed integer program [4,19]

$$\begin{aligned} \text{minimize} \quad & \sum_{t=1}^{T}\sum_{k=1}^{K} \ell\Big(y(t) - \hat{\theta}_k^{\mathrm{T}}\Phi(t)\Big)\mathcal{X}_{t,k} \\ s.t. \quad & \sum_{k=1}^{K}\mathcal{X}_{t,k} = 1 \ \ \forall t \in [1 \ldots T] \\ & \mathcal{X}_{t,k} \in \{0, 1\}, \end{aligned} \tag{7}$$

where $\ell(\cdot)$ is an arbitrary p-norm. The minimization is with respect to the parameter estimates $\hat{\theta}_k$, $k = 1 \ldots K$ and the matrix $\mathcal{X}$, consisting of $T \times K$ binary variables. The significance of the binary variable $\mathcal{X}_{t,k}$ is to indicate whether or not a data point belongs to submodel k at time instance t. The discrete state can be recovered as

$$\eta(t) = k \iff \mathcal{X}_{t,k} = 1. \tag{8}$$

Note however that solving this kind of problems require exhaustive search algorithms with worst-case complexity that is exponential in the size of the input, $\mathcal{O}(K^T)$. Practically, this means enumerating all the possible integer combinations of the discrete state. Therefore the mixed integer programming approach is computationally affordable only for problems with short time-series [17,19].

Other approaches involve heuristics in order to deliver estimates in tractable time, with the trade-off that the estimates are only approximate solutions. In this work, SON-EM, a three-step algorithm introduced by Hartmann et al. [12] was used for system identification and will be briefly described here.

As a preliminary step, a Sum Of Norm (SON) regularized least squares optimization problem is solved

$$\text{minimize} \ \sum_{t=1}^{T} ||y(t) - \tilde{\theta}^{\mathrm{T}}(t)\Phi(t)||^2 + \lambda \sum_{t=2}^{T} ||\tilde{\theta}(t) - \tilde{\theta}(t-1)||_p \,. \tag{9}$$

Here the optimization is with respect to $\tilde{\theta}(t)$, yielding the parameter estimates after the first step. The optimization problem in Eq. (9) implies a positive scalar regularization parameter, λ, which is the only one tuning parameter of the method. Reasonable λ values are typically of magnitude between 0.1 and 10, However factors like the signal to noise ratio, and the values of the parameters may influence the ideal choice of λ. It was found in [12] on a slowly varying system, that the final results are not highly sensitive to the tuning parameter. This is a major advantage of the identification method, because the tuning parameter does not have to be exactly determined in order to yield good final results. In the rest of this study, for all the experimental data, the tuning parameter was set uniformly to $\lambda = 0.5$.

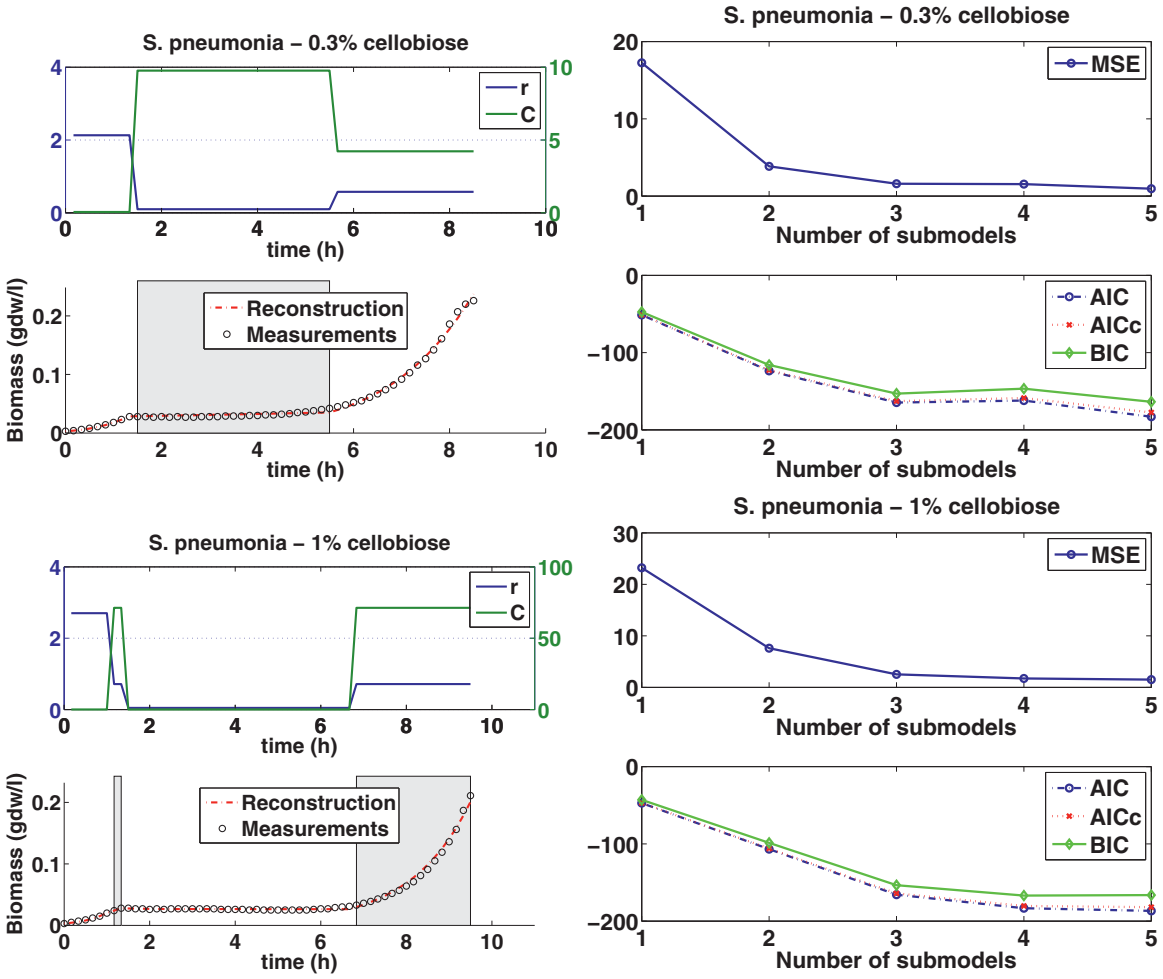

Fig. 3. *Streptococcus pneumoniae* D39 grown on glucose and cellobiose as presented by Boianelli et al. [5,6].

The expression $||\cdot||_p$ denotes for the p-norm, where the design variable p plays a crucial role by setting the norm of the regularization. Imposing a smoothing condition on the switches of the parameters has computationally convenient properties since, for $p \geq 1$, one obtains a continuous convex problem. If sparsity of the solution is favored, in the sense that the change of one dimension in the parameter vector is preferred against change in all the dimensions, p should be set to one, leading to a least absolute shrinkage and selection operator (Lasso) flavor ℓ_1-norm regularization, similar to group Lasso [23].

Although the solution of Eq. (9) is an estimate of the parameters, since here the cardinality of the submodel set is not limited to K, it is not guaranteed to be a feasible point of the original problem. The rationale to apply Expectation Maximization (EM) clustering for mixture of Gaussians as described in details in [12, Appendix A], in the second step is to constrain the cardinality of the submodel set and to estimate the switching sequence. Note that theoretically any other unsupervised learning algorithm could be applied here, that casts the preliminary estimates into K distinct sets, such as clustering or labeling methods.

The solution after the second step is feasible, but may not be optimal in the mean squared error sense because the clustering only operates on $\tilde{\theta}(t)$, independently from the observations. This problem is tackled in the third step by replacing the estimated switching sequence into Eq. (7) such as to obtain the following convex quadratic problem

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{t=1}^{T} ||y(t) - \hat{y}(t)||^2 \\ s.t. \quad & \hat{y}(t) = \hat{\theta}_{\hat{\eta}(t)}^{\mathrm{T}} \Phi(t)\,. \end{aligned} \tag{10}$$

Here the optimization is only with respect to $\hat{\theta}_k$, $k = 1 \ldots K$, because $\hat{\eta}(t)$ is given from step 2. Solving this optimization problem in the third step yields parameter estimates that are optimal according to the estimated switching sequence. The SON-EM algorithm is summarized in Algorithm 1.

Algorithm 1 SON-EM algorithm for parameter estimation of hybrid time-varying parameter systems.

1. Solve the convex optimization problem of Eq. (9) to deliver preliminary estimates $\tilde{\theta}$
2. Proceed with EM clustering on $\tilde{\theta}$ to obtain K classes
3. Solve the convex optimization problem of Eq. (10)

Besides fitting the model, it is important to estimate the appropriate number of subsystems. The biological analysis of the processes for *L. lactis* and *S. pneumoniae* suggest three phases, namely an initial growth phase, a lag phase and a secondary growth phase. Accordingly, the number of submodels is expected to be $K = 3$. For *S. cerevisiae*, the growth rate was reported to decrease at the end of the first growth phase [21, Supplement], thus here the number of submodel is supposed to be $K = 4$. To support (or contradict) these hypotheses, model selection was performed, where the estimation is performed in an iterative process as follows. Fittings are made

Table 1
Multiphasic growth datasets.

Organism	Strain	Reference
Lactococcus lactis	NZ9000	This article
Streptococcus pneumoniae	DP1004	[5,6]
Saccharomyces cerevisiae	Wild type	[21]

with increasing number of subsystems, and the Mean Squared Error (MSE) of each reconstruction is computed. As a rule of thumb, the MSE decreases with increasing K, the optimal value is reached when the change in MSE is sufficiently small. This is justified by the fact that if no significant change in MSE is observed, then introducing a new subsystem would not further improve the reconstruction. Also, the common measures for model selections such as Akaike Information Criterion (AIC), AIC corrected for small sample size (AICc) and Bayesian Information Criterion (BIC) [1,7] were computed.

$$\mathrm{AIC} = T \ln\left(\widehat{\sigma_\varepsilon^2}\right) + 2K \tag{11}$$

$$\mathrm{BIC} = T \ln\left(\widehat{\sigma_\varepsilon^2}\right) + K \ln(T) \tag{12}$$

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2K(K+1)}{T-K-1} \tag{13}$$

where $\widehat{\sigma_\varepsilon^2}$ is the empirical error variance

$$\widehat{\sigma_\varepsilon^2} = \frac{1}{T}\sum_{t=1}^{T}\left(x(t) - \hat{x}(t)\right)^2 \tag{14}$$

2.3. Dataset

Growth data of three different organisms were used as presented in Table 1. *Lactococcus lactis* strains NZ9000[pGalPMKT] and NZ9000[pGalPMKTpgmA] [16] were grown in batch mode in chemically defined medium containing 1% (w/v) galactose under anaerobic conditions in rubber-stoppered bottles (200-ml) statically at 30 °C. Chloramphenicol was used at 5 μg ml^{-1}. Expression of genes cloned downstream of the nisin-inducible P_{nisA} promoter was induced when the optical density at 600 nm (OD_{600}) was 0.25 by addition of a nisin solution (1 μgl^{-1} in a 50% (v/v) ethanol solution). Prior to nisin addition cells were growing exponentially at maximal growth rate of 0.31 h^{-1}, a value slightly lower than that observed for the strain NZ9000 harboring the empty vector (μ, 0.34 h^{-1}). Addition of nisin caused severe growth impairment for a period of 12 h (between the 5th and the 17th hour), followed by a second growth phase at a maximal rate of 0.19 h^{-1}. The biphasic growth behavior can be derived from a metabolic burden due to overexpression of four proteins or a metabolic imbalance caused by accumulation of a toxic intermediate. It was also observed that lowering the concentration of nisin added to cells growing on galactose attenuates the biphasic profile (data not shown). It should be noted, however, that glucose-grown cells do not display biphasic growth upon nisin-induction of the *nisA* promoter. Interestingly, nisin-induction of the *nisA* was reported to be significantly impeded in the presence of galactose and an active Leloir pathway [9]. These authors proposed that galactose competes with nisin for the same or overlapping binding sites on the *nisA* promoter, and thus it is not implausible to suggest that this antagonistic effect leads to a transient growth defect.

The *Streptococcus pneumoniae* data was originally published in [5,6]. Fittings were made to two time-series with growth on glucose and beta-glucoside cellobiose, where the cellobiose concentration was of 0.3% and 1%, respectively.

Wild type *Saccharomyces cerevisiae* was grown in batch conditions on minimal medium containing 11.11 mM glucose as a sole source of carbon and energy as described in [21].

3. Results

Almost perfect reconstructions were achieved to all the time-series, as simulation overlaps with experimental data when estimating with $K = 3$ submodels. For the *L. lactis* data, the model selection shown in Fig. 2 similar MSE scores for two and three subsystems were obtained, just slightly higher than for four subsystems. Thus, from a computational point of view, the two-subsystem model should be preferred. On the other hand, biological insights into the process suggest three subsystems: an initial growth phase, a growth impairment period (lag), and a secondary growth phase. Thus, in order to compare with *a priori* biological knowledge, the three-subsystem model was chosen for further analysis. We also report the results for a two-subsystem model (see reconstruction in Appendix A, Fig. A.5.), which could also have a biological interpretation: after an initial growth impairment period, the growth is resumed with the same parameters as in the initial growth phase. This would be an interesting hypothesis to test experimentally. The time of transition to the lag phase was well identified, however the length of the detected lag phase was longer than the ones determined by the experts (at 19 h instead of 14 h). Table 2 shows that the automatically identified maximal growth parameters are close, but slightly above the parameters identified by hand.

The model yielded a good fit to the *Streptococcus pneumoniae* data as seen in Fig. 3. Model selection predicted the number of subsystems to be three for both experiments, corresponding to the first growth phase, the lag and the second growth phase. For lower cellobiose concentration the lag phase is significantly shorter or inexistent.

Table 2
Estimated parameters for multiphasic growth datasets. The parameters are grouped by the species. Values determined by the experts are in parenthesis when available.

			θ_1	θ_2	θ_3	θ_4
L. lactis	pGalPMKT	**r**	0.34 (0.31)	0.02 (0.02)	0.21 (0.19)	
		C	33.87	2.20	2.96	
	pGalPMKTpgmA	**r**	0.37 (0.33)	0.07 (0.03)	0.29 (0.21)	
		C	36.52	0.73	2.51	
S. pneumoniae	0.3% cellobiose	**r**	2.13	0.10	0.58	
		C	0.05	9.76	4.22	
	1% cellobiose	**r**	2.70	0.05	0.71	
		C	0.03	0.05	71.26	
S. cerevisiae		**r**	0.38 (0.40)	0.87	0.21	0.17 (0.05)
		C	2.23	1.74	1.36	5.26

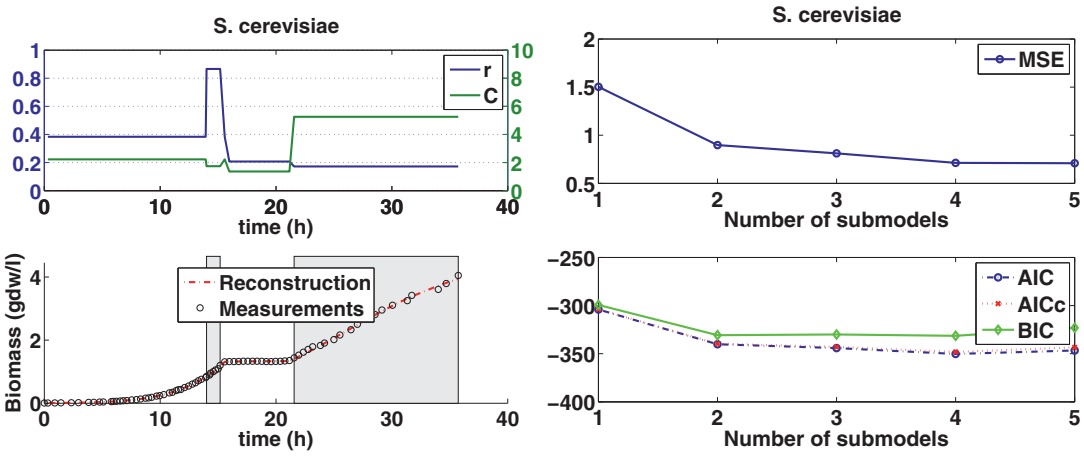

Fig. 4. Growth of *Saccharomyces cerevisiae* as described by Slavov et al. [21].

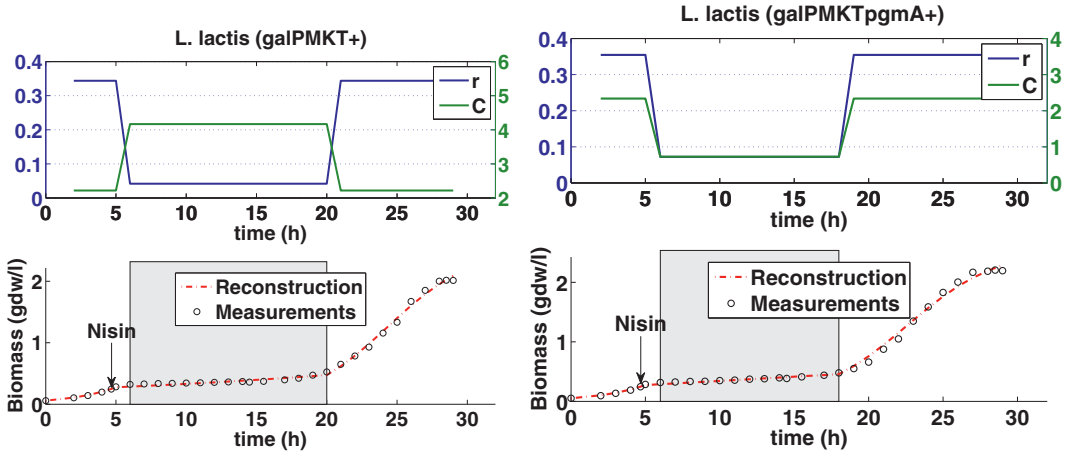

Fig. A.5. *Lactococcus lactis* growth on galactose. Expression of galPMKT under the nisin promoter was induced by addition of nisin (1 µg/l) at OD600 of 0.25.

Regarding the results obtained for *S. saccharomyces*, the expected number of phases is three: a glucose growth phase a lag without growth, and an ethanol growth phase. However growth rate was reported to be declined at the end of the first growth phase [21, Supplement]. This corresponds to the model selection results in Fig. 4 indicating 4 phases. The automatic segmentation shows a good match to previously published results. Estimated growth parameters for the glucose phase are close to the reported ones, while the estimated growth rate on the ethanol phase remains higher than the reported value. The reason behind this difference is presumably that previously fitting was only made to the second part of the last phase. The complete results for all the three datasets using one to five subsystems are presented in Supplementary material.

4. Discussion

A hybrid model for multiphasic bacterial growth has been presented, including $2 \times K$ parameters, where K is the number of subsystems considered. Unlike recently developed models [5,6,8,20], the proposed model is linear, which also follows that there are no identifiability issues. The simplicity of the model originates from its switching nature. The separate phases of the process are modeled with the same dynamics, but with different parameters for each phase. The fact that in all the cases the model could explain the experimental data illustrates its relevance.

It is arguable that the proposed model does not incorporate parameters of the cellular mechanisms of the organism, and as such, its predicting ability is limited. On the other hand, in the conventional setups the only available measurement is biomass data, and parameters of more complicated models are either interpolated or parameters from historical studies are used, moreover, they may suffer from identifiability issues. The hybrid interpretation of the process is that the organism shows constant behavior (parameters do not change) between the discrete events (e.g. the collapse of a carbon source or overcoming of a toxic effect) indicating the changes of parameters. And however the current modeling framework does not allow the prediction of the discrete events, analysis of retrospective data is promising.

The obtained segmentation results were close to the ones determined by the experts, however for the *L. lactis* data the number of subsystems was estimated to be two against the expected three, although this alternative hypothesis can also be biologically interpretable. Moreover the length of the detected lag phase for the *L. lactis* data was slightly longer then determined by the experts. As the growth phase itself may start with a lag phase, in general it is hard to decide the exact time of the transition when the dynamics is low. To identify logistic models of longer lag times, it is suggested to make the fittings to the logarithmic growth ratio [18,24]. As in the data presented, most of the growth phases are not expected to start with lag, instead of applying the logarithmic transformation, lags were considered as separate phases. Other bacterial growth models like the Gompertz [24] or Baranyi model [3] may

provide more flexibility, and will be addressed in future research.

5. Conclusion

A hybrid SARX model for multiphasic bacterial growth has been presented, and analyzed in the light of real experimental growth data from three different species. The datasets were segmented and parameters were identified using the SON-EM method [12]. The results indicate that in most cases the partitioning of the data corresponds to what was determined by the experts, and the identified growth parameters are also close to the published ones. An overall very good reconstruction to the experimental dataset was achieved on all the species showing the adequacy of the modeling approach. It is hypothesized that this could be a new automatic way to partition and identify corresponding parameters of experimental multiphasic growth data or other processes consisting of multiple linear phases.

Acknowledgments

The authors thank to Nikolai Slavov, Marco R. Oggioni and Alessandro Bidossi for permission to access the growth datasets. This work was partially supported by national funds through FCT-Fundação para a Ciência e a Tecnologia under the project LAETA 2015–2020, reference UID/EMS/50022/2013, project INESC-ID UID/CEC/5021/2013 supported by FCT, project PERSEIDS (PTDC/EMS-SIS/0642/2014), and by European Union Framework Program 7 "BacHBerry" (www.bachberry.eu), Project no. FP7-613793. AH received a PhD fellowship from FCT Portugal under the reference SFRH/BD/69336/2010. SV acknowledges support by Program Investigador FCT (IF/00653/2012), co-funded by the European Social Fund through the Operational Program Human Potential.

Appendix A. Reconstruction of *L. lactis* with two subsystems

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.mbs.2016.06.013

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.
[2] A. Amrane, L. Adour, C. Couriol, An unstructured model for the diauxic growth of Penicillium camembertii on glucose and arginine, Biochem. Eng. J. 24 (2005) 125–133.
[3] J. Baranyi, T.a. Roberts, A dynamic approach to predicting bacterial growth in food., Int. J. Food Microbiol. 23 (3–4) (1994) 277–294.
[4] A. Bemporad, J. Roll, L. Ljung, Identification of hybrid systems via mixed-integer programming, in: IEEE Conference on Decision and Control, Orlando, FL, USA, 2001, pp. 786–792.
[5] A. Boianelli, A. Bidossi, L. Gualdi, L. Mulas, C. Mocenni, G. Pozzi, A. Vicino, M.R. Oggioni, A non-linear deterministic model for regulation of diauxic lag on cellobiose by the pneumococcal multidomain transcriptional regulator CelR., PloS ONE 7 (10) (2012).
[6] A. Boianelli, A. Bidossi, C. Mocenni, L. Mulas, G. Pozzi, M.R. Oggioni, A. Vicino, A nonlinear dynamic model for diauxic growth of Streptococcus pneumoniae on glucose and cellobiose, in: IFAC World Congress, Milano, Italy, 2011, pp. 11773–11778.
[7] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference - A practical Information-theoretic approach, vol. 40, 2nd, Springer, 2002.
[8] A.I. Casasús, R.K. Hamilton, S.A. Svoronos, B. Koopman, A simple model for diauxic growth of denitrifying bacteria, Water Res. 39 (9) (2005) 1914–1920.
[9] S. Chandrapati, D.J. O'Sullivan, Characterization of the promoter regions involved in galactose- and nisin-mediated induction of the nisA gene in Lactococcus lactis ATCC 11454, Mol. Microbiol. 46 (2) (2002) 467–477.
[10] L. Galdieri, S. Mehrotra, S. Yu, A. Vancura, Transcriptional regulation in yeast during diauxic shift and stationary phase., Omics : J. Integr. Biol. 14 (6) (2010) 629–638.
[11] A. Garulli, S. Paoletti, A. Vicino, A survey on switched and piecewise affine system identification, in: 16th IFAC Symposium on System Identification, 2012 Brussels, Belgium, 2012, pp. 344–355.
[12] A. Hartmann, J.a.M. Lemos, R.S. Costa, J.a. Xavier, S. Vinga, Identification of switched ARX models via convex optimization and expectation maximization, J. Process Control 28 (2015) 9–16.
[13] J. Lygeros, Lecture notes on hybrid systems, in: Notes for an ENSIETA workshop, 2004.
[14] J. Monod, Recherches sur la croissance des cultures bacteriennes, 1942 Ph.D. thesis.
[15] J. Monod, The growth of bacterial cultures, Ann. Rev. Microbiol. 3 (1949) 371–394.
[16] A.R. Neves, W.a. Pool, A. Solopova, J. Kok, H. Santos, O.P. Kuipers, Towards enhanced galactose utilization by Lactococcus lactis., Appl. Environ. Microbiol. 76 (21) (2010) 7048–7060.
[17] S. Paoletti, A.L. Juloski, G. Ferrari-Trecate, R. Vidal, Identification of hybrid systems - a tutorial, Eur. J. Control 13 (2–3) (2007) 242–260.
[18] M. Peleg, M.G. Corradini, M.D. Normand, The logistic (Verhulst) model for sigmoid microbial growth curves revisited, Food Res. Int. 40 (7) (2007) 808–818.
[19] J. Roll, A. Bemporad, L. Ljung, Identification of piecewise affine systems via mixed-integer programming, Automatica 40 (2004) 37–50.
[20] M. Santillán, M.C. Mackey, Quantitative approaches to the study of bistability in the lac operon of Escherichia coli., J. R. Soc. Interface 5 (Suppl 1) (2008) 29–39.
[21] N. Slavov, B.A. Budnik, D. Schwab, E.M. Airoldi, A. van Oudenaarden, Constant growth rate can be supported by decreasing energy flux and increasing aerobic glycolysis., Cell Rep. 7 (3) (2014) 705–714.
[22] A. Veríssimo, L. Paixão, A.R. Neves, S. Vinga, BGFit: management and automated fitting of biological growth curves, BMC Bioinform. 14 (1) (2013) 283.
[23] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc.: Ser. B 68 (2006) 49–67.
[24] M.H. Zwietering, I. Jongenburger, F.M. Rombouts, v.K. Riet, Modeling of the bacterial growth curve, Appl. Environ. Microbiol. 56 (6) (1990) 1875–1881.